STIC-ILL

From: Schmidt, Mary
Sent: Thursday, May 16, 2002 2:41 PM
To: STIC-ILL
Subject: references, 09/646,825

please locate the following:

1. Bau et al, NAR 22:2811-2816 (1994)
2. Dean et al., NAR 14:2229-2240 (1986)
3. Malter et al, Science 246:664-666 (1989)
4. Shaw et al., Cell 46:659-667 (1986)
5. van Aarsen et al., Plant mol. biol. 28:513-524 (1995)
6. Wilson et al., Nature336:396-399 (1998).

QH 433. P5

Thanks,
Melissa
11D05
11th floor mailbox

Plant Molecular Biology **28**: 513–524, 1995.

cry IA(b) transcript formation in tobacco is inefficient

Roel van Aarssen, Piet Soetaert, Maike Stam[1], Jan Dockx[2], Veronique Gosselé, Jef Seurinck, Arlette Reynaerts and Marc Cornelissen*
Plant Genetic Systems NV, J. Plateaustraat 22, B9000 Gent, Belgium (author for correspondence);*
[1] present address: Dept. of Genetics, Free University, De Boelelaan 1087, Amsterdam 1081 HV, Netherlands;
[2] present address: Dept. of Molecular and Cellular Biology, Utrecht University, Padualaan 8, Utrecht 3584 CH, Netherlands

Received 11 July 1994; accepted in revised form 13 April 1995

Key words: *cry* genes, mRNA stability, RNA splicing

Abstract

Chimaeric $P_{CaMV35S}$*cry* genes direct in tobacco mesophyll protoplasts mRNA levels of less than one transcript per cell. We provide evidence that this low cytoplasmic *cry* IA(b) mRNA level is not due to a rapid turnover but rather results from a marginal import flow of *cry* messenger into the cytoplasm. Run-on assays indicate that the frequency of transcription initiation is not limiting. However, the *cry* precursor mRNA carries at least three regions that are recognized as introns. The absence of high cytoplasmic levels of spliced *cry* mRNAs suggests that these mRNAs are unstable and/or not efficiently made. Point mutations in the 5′ splice site of the most distal intron allows high accumulation levels of the full-length mRNA. This implies that the inefficient formation of full-size mRNA is a major cause of the low expression level of chimaeric *cry* IA(b) genes in tobacco.

Introduction

The regulatory features of DNA and RNA sequences are mostly context dependent. Local context changes thus can affect gene regulation in a significant and yet unpredictable manner. In this way, transgenes may follow unexpected expression pathways, both because of cryptic regulatory signals and because of changes in the genetic background.

A well documented case of this sort is the expression of insecticidal crystal protein (*cry*) genes in plants. *Cry* genes are naturally present in *Bacillus thuringiensis* and encode proteins which are lethal to specific insect species upon ingestion [20]. Plants expressing the appropriate *cry* gene become significantly less damaged by specific insect infestations than plants lacking the gene. Transgenic tobacco, tomato, potato and cotton carrying a number of different *cry* genes specified without exception barely detectable levels of insecticidal crystal protein (ICP) [34, 41, 42]. Typically, chimaeric *cry* plant genes direct about 1000-fold lower levels of protein than chimaeric *bar* (bialophos resistance), *nptII* (neomycin phosphotransferase), *cat* (chloramphenicol acetyltransferase), or *gus* (glucuronidase) genes even though being flanked by the same plant regulatory expression signals [7, and unpublished]. RNA analyses [42] and protein turnover studies (Cornelissen and Vandewiele, unpublished) revealed that the minimal ICP accumulation is primarily

due to low cytoplasmic *cry* IA (b) mRNA steady state levels. Introduction of a large number of translationally neutral mutations in the *cry* IA(b)-coding region resulted in a significant increase of the mRNA and protein level [35]. However, the underlying mechanism restricting *cry* expression is still not understood.

In the present study we have focused on the question of which events negatively control *cry* IA(b) expression in tobacco. We have determined the turnover rate of the *cry* messenger and provide evidence that the level of cytoplasmic import rather than mRNA instability is responsible for the inefficient cytoplasmic accumulation of *cry* mRNA. Run-on analysis suggests that the promoter functions normally. RT-PCR analyses of *cry* transcripts revealed the occurrence of three splicing events. Mutation of the 5′ splice sites showed that cryptic splicing of the *cry* messenger contributes significantly to the low *cry* transcript flow to the cytoplasm.

Materials and methods

DNA constructions

The *in vitro* transcription vectors used to produce *bar* and *bt884* RNA were derived from a modified pFM108 [10] which is a pGEM (Promega) derivative containing a cloning site at the T7 transcription start site, the pGEM1 polylinker and a $(A)_{33}G(A)_{32}G(A)_{32}$ sequence followed by respectively a *Kpn* I (*bt884*) or *Hin*d III (*bar*) site (pMS vector, Maike Stam, thesis, Free University of Amsterdam, Amsterdam, 1991). The *bar*-coding region was introduced into the *Nco* I and *Bgl* II sites of pMS yielding pXD324. The *bt884* sequence was introduced into the pMS vector by introduction of the 1840 bp *Bam* HI (filled) – *Pst* I (trimmed) *cry* IA(b) fragment into the *Nco* I (filled)-*Eco* RI (filled) sites of pMS resulting in pMS-IA(b). The presence of the poly(AT) track and the flanking regions were sequence-analysed [30]. The poly(AT) track in pMS-IA(b) had changed into $(A)_{33}G(A)_{32}G(A)_{21}$.

The chimaeric $P_{TR2'}$IA(b)-6 3′*g7* gene is a derivative of the $P_{TR2'}$*bt884* 3′*g7* gene and contains a consensus translation start context [23], a 5′ truncation of the coding region from codons 2 to 28, as well as a modified 321 bp *Xba* I-*Eco* RI fragment (nucleotides 674 to 1000) relative to the AUG of the *bt884*-coding region (Fig. 5a). The IA(b)-14 gene is a derivative of IA(b)-6 in which the 5′ splice site of intron 2 is modified from 5′-CGG′ gtaaga into CGTGTCCGA by PCR methodology [22]. The IA(b)-16 gene is a derivative of IA(b)-14 in which the 5′ splice site of intron 1 is modified from 5′AGT′gtaagt into TCTGTTTCG by overlap extension PCR [22]. The presence of the modifications in the selected IA(b)-14 and -16 clones was confirmed by sequence analysis [36]. For transient expression analyses the chimaeric IA(b)-6, -14 and -16 genes were introduced into a pUC19 [43] derivative upstream of chimaeric P_{35S}*cat* 3′*ocs* gene in a head-to-tail configuration, resulting in respectively plasmids pPS0212, pRVA0208 and pRVA0210. For stable introduction into tobacco the chimaeric IA(b)-6 gene was exchanged with the chimaeric *bt884* gene present in the T-DNA vector pGSH163 [42] resulting in pPS0216.

Tobacco transformation

The T-DNAs of plasmids pGSH163 and pPS0216 containing respectively the *bt884* and IA(b)-6 gene were introduced into *Nicotiana tabacum* cv. Petit Havanna SR1 [28] via *Agrobacterium*-mediated T-DNA transfer as described by Vaeck *et al.* [42].

Determination of RNA turnover and transcript analyses

pMSIA(b) and pXD324 were linearized at the unique *Kpn* I or *Hin*d III site downstream the poly(AT) track. Synthetic capped RNA was produced according to Promega protocols. The use of the linearization sites *Kpn* I and *Hin*d III resulted in the production of transcripts with 2 and 3 nucleotides respectively downstream of the

poly(A) tail. A limited number of nucleotides downstream of the poly(A) tail appears not to interfere with the stability of transcripts [14]. Free nucleotides were removed by two ethanol precipitations in the presence of 2 M ammonium acetate. RNA was delivered in mesophyll protoplasts of *N. tabacum* cv. Petit Havanna SR1 [28] essentially as described by Gallie *et al.* [14] using the buffer systems as in Denecke *et al.* [11]. 9 pmol of synthetic capped and polyadenylated *cry* and *bar* transcript were mixed and added to 10^6 protoplasts in 0.3 ml of electroporation buffer. After electrical discharge and a 10 min incubation at room temperature, the cells were pooled and washed in 10 volumes of buffer (0.2 M mannitol, 128 mM NaCl, 4 mM $CaCl_2$, 0.15 mM KH_2PO_4, 0.81 mM Na_2HPO_4, pH 7.2), resuspended in culture medium and incubated at 24 °C in the dark until harvesting. Time point samples contained 3×10^6 protoplasts.

Total RNA was extracted from tobacco protoplasts as described by Cornelissen and Vandewiele [8]. Equal amounts (1–10 μg) of different samples were run on a denaturing gel. To quantify the *cry*, *bar* and *nptII* mRNA levels in the total RNA preparations, a dilution series ranging from 1–500 pg of synthetic *cry*, *bar* and *nptII* transcript were run on the same gel. To each dilution, an identical quantity of total SR1 RNA was added as was present in the time point samples. Filters and hybridizations were done according to Amersham protocols. Probes were made by *in vitro* transcription of pGEM clones that carried the *cry*, *bar* and *nptII* coding sequences. Hybridisation signals were quantified by scintillation counting and densitometric scanning using a LKB Ultrascan 2202. Data were processed using GraphPAD Inplot (3.0) software.

cDNA reactions to determine *cry* IA(b) RNA processing sites were carried out according to Kawasaki [24] on 2 μg of total RNA extracted from N28-220 [42] leaf protoplasts using an oligo d(T) primer of 16 nt. The first strand was used to amplify specific DNA fragments employing the Polymerase Chain Reaction (PCR) as described by Kawasaki [24] using the following sets of primers: set A: OPS64 (5′-ATC CGG TCC CCA TAC ACG CTC TAA TCC CG) and OPS66 (5′-CCA AAT CGA TGG ATC CCG ATA ACA ATC CG), set B: OPS63 (5′-TTG GAC AAA GGT GGG GAT TTG ATG CCG GG) and OPS65 (5′-ATG GCT TAA TCG ATG ACT AAA TCC TTG CC), set C: OPS58 (5′-GTG CCA CCT AGG CAA GGA TTT AGT CAT CG) and OPS59 (5′-TAG CTC ATC GGG GGA TCT GCT AGA GCC CG), and set D: OPS61 (5′-ACA TTG CCG TAG ATG AAA GAC TGA GTG CG) and OPS62 (5′CCC CAA TCG ATA TTT CCT TGT CGC TAA CG). The amplified DNA products were separated on agarose gel, purified and sequenced according to Maxam and Gilbert [30] or Sanger *et al.* [36]. Agarose gel separation of PCR samples resulted in more PCR species than polyacrylamide gel separation. Purification and sequence analysis of both strands of the additional PCR species showed that they originated from heteroduplex formation [12, 46]. cDNA synthesis and PCR to evaluate the splice site modifications were performed using the 3′-AmpliFINDER RACE Kit protocol (Clontech Laboratories). Primer OPS60 (5′-ACG GTA GAT TCG CTG GAT GAA ATA CCG CC) was used in the primary PCR, primers OPS58 and OPS59 were utilised in the secondary PCR. The PCR products were analysed on polyacrylamide gel.

Nuclear run-on assays

Preparation of transcriptionally active nuclei and run-on assays were done essentially as described by Cox and Goldberg [9]. Assays were performed for 20 min. The radioactively labelled RNA was extracted according to Celano *et al.* [5]. The Southern (Hybond N of Amersham) shown in Fig. 2 contained 0.2 pmol of the denatured DNA fragments. Negative control DNAs were included. Hybridizations were carried out at 42 °C in 50% formamide, 5× SSPE and 0.5% SDS essentially as described in Amersham protocols. The autoradiographs were scanned with a LKB Ultrascan 2202 and normalized as described in the legend to Fig. 2.

DNA delivery into tobacco protoplasts

DNA was delivered into SR1 mesophyll protoplasts as described by Denecke *et al.* [11]. Samples for RNA analysis were taken four hours after DNA delivery. Samples were processed as described for the RNA turnover studies. Equal amounts of each sample were run in duplicate on denaturing formaldehyde gels and blotted. Dilution series of synthetic *cry* and *cat* RNA were present on the same northern blot to quantify the respective mRNA levels. The probes were made by *in vitro* transcription of pGEM derivatives carrying the *bt884*- and *cat*-coding sequences. Transcript levels were quantified by densitometric scanning (LKB Ultrascan 2202) and scintillation counting of the hybridization signals.

Results

cry *messengers are stable transcripts in tobacco*

Transgenic tobacco plant N28-220 [42] which carries several copies of the *cry* IA(b) derived $P_{TR2'}$-*bt884*-3′*g7* gene, specifies in its leaves less than one *bt884* transcript per cell. N28-220 is amongst the lines that express the highest level of ICP. Its *cry* transcript level is over 200-fold below that of the *bar* transcript in tobacco SR1(T-GSFR166) [7], which carries the *bar* coding sequence flanked by the same regulatory sequences as the *cry* gene. Similar observations were made with transgenic tobacco expressing chimaeric *cry* IB [3], *cry* IC [21] and *cry* IE [2] genes (data not shown).

In a first instance, we wished to understand to what extent nuclear and/or cytoplasmic events determine the unusually low cytoplasmic *bt884* mRNA level. This distinction can be made on the basis of the half-life and steady-state level of the *cry* transcript as cytoplasmic mRNA levels are a function of import and turnover. At equilibrium the cytoplasmic steady-state level equals the import rate divided by the decay constant.

The steady-state *bt884* transcript level in leaf protoplasts of N28-220 is too low to determine the turnover rate by pulse-chase analyses [26]. We therefore determined the turnover of the *cry* transcript in two independent ways. First, the *in-vivo* stability of *in vitro* synthesized *bt884* transcript was determined [13, 14, 15]. A T7 template was constructed to synthesize transcripts with a 5′ cap and a 3′ poly(A) track carrying the coding sequences of *bt884* and *bar*. At different time points after delivery samples were taken and analysed for both the *cry* and *bar* transcript abundances. As shown in Fig. 1A and 1B the *cry* transcript decays at a rate similar to that of the *bar* transcript. Using GraphPad Prism (1.02) software a decay constant of 0.088 ± 0.038/h and 0.14 ± 0.026/h was calculated for the *cry* and *bar* transcripts respectively. The biologicial significance of these values is illustrated by the similar decay rates of synthetic *bar* mRNA and an analogous gene-encoded *bar* mRNA. This latter decay rate was obtained after transcription inhibition of a stably integrated transgene [7] (decay constant 0.3/h). As a second approach to determine the cry transcript stability we incubated mesophyll protoplasts of N28-220 with actinomycin D (150 μg/ml) and quantitated the amount of bt884 mRNA at regular time intervals over a period of 6 h (not shown). This experiment showed a decay constant of 0.13 ± 0.034/h for the *bt884* mRNA.

These results show that the *cry* IA(b)-coding sequence itself is not subject to any specific decay process upon entry in the cytoplasm. Therefore, in plants like N28-220, *bt884* messenger could accumulate in principle as efficiently as the *bar* messenger in SR1(T-GSFR166). As such accumulation does not take place, it implies that *cry* mRNA enters the cytoplasm at very low rates. Based on a steady state level of about one transcript per cell, the nucleo-cytoplasmic flow rate of *bt884* mRNA should be less than 1 transcript per 10 h (import rate = steady state × decay constant). This clearly points towards the occurrence of one or more nuclear events that impair *cry* expression.

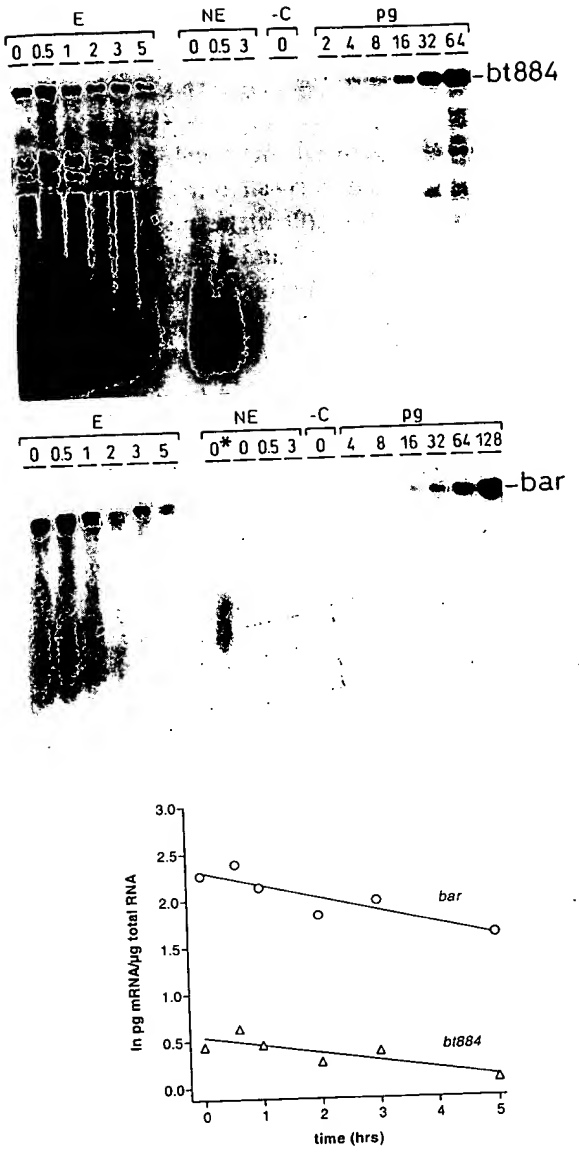

Fig. 1. Determination of the cytoplasmic stability of *bt884* transcript in tobacco mesophyll protoplasts. A (top) E = Electroporated protoplasts. *In vitro* transcribed, capped and polyadenylated *bt884* and *bar* transcripts were introduced into SR1 protoplasts and total RNA was extracted at the time points indicated (h). Time point 0 is 40 min after delivery. Transcript levels were quantified with the aid of calibration curves obtained with dilution series of *cry* and *bar* transcripts present on the northern blots as described in Materials and methods. The same *cry* and *bar* RNA prepararions were used for the delivery and dilution series. Degradation products visible in the lower part of the (E) and (NE) samples result from protoplast-encoded RNAse activities. The introduced quantity of *bar* mRNA (60 pg/µg total RNA) is similar to that accumulating in protoplasts containing a stably introduced *bar* transgene [7]. NE = not electroporated protoplasts: intact extracellular RNA was undetectable within 1 h after RNA addition, in protoplast samples that were mixed with *cry* and *bar* transcripts and that were not subjected to an electrical discharge. C = protoplasts electroporated in the absence of RNA. In all lanes 2.5 µg of total RNA was loaded apart for the lane (NE)-0* to which 5 µg was applied. B (bottom). The calculated half-lives for *bt884* and *bar* mRNA are 7.8 ± 3 and 4.9 ± 1.3 h respectively.

The bt884*-coding region does not interfere with transcription initiation*

It has been shown that sequences downstream of the transcription start site can be involved in control of transcription initiation [27]. We, therefore, investigated whether the *cry*-coding sequence interferes with the activity of the *TR2'* promoter. If this were the case, the number of RNA polymerase II complexes actively transcribing the *bt884*-coding region would be relatively low. In order to have a reference, we compared the transcriptional activity on the *nptII*- and *bt884*-coding region in N28-220. The *nptII* and *bt884* genes are under the transcriptional control of respectively the *TR1'* and *TR2'* promoter and the *nptII* and *bt884* transcript levels differ over 50-fold (Fig. 4b). We prepared nuclei of N28-220 mesophyll protoplasts and performed a nuclear run-on assay in the presence of radioactively labelled ribonucleotides. After the extension reaction, the labelled RNA was extracted and allowed to hybridize with a Southern blot that contained a molar excess of denatured DNA fragments covering both the *nptII*- and *bt884*-coding region (Fig. 2). Quantitation of the amount of radioactively labelled RNA hybridizing with the *nptII* and *bt884* DNA fragments revealed that the transcriptional activity directly downstream of the promoter region is approximately the same for both genes. This result strongly suggests that the main nuclear mechanism by which the *bt884*-coding region interferes with *cry* expression is not a down regulation of the number of transcription initiations but rather a more downstream event.

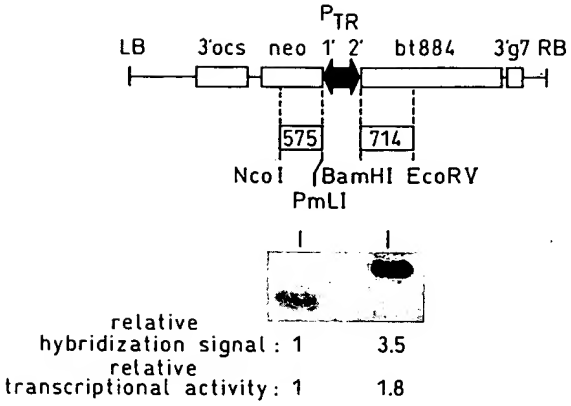

Fig. 2. Comparison of the transcriptional activity of *neo* and *bt884* genes near the transcription start site in nuclei of N28-220. The RNA extension reaction was performed in the presence of ^{32}P-UTP for 20 min at 30 °C. Radioactively labelled RNA extracted from nuclei was hybridised with a molar excess of denatured *neo* and *bt884* DNA which was immobilized on a nylon membrane. The positions and sizes of the DNA regions on the T-DNA are indicated. LB and RB refer to the left and right T-DNA borders, respectively. The 3′ end formation and polyadenylation region of the *g7* and *ocs* genes [42] are indicated. The relative abundance of hybridising *neo* and *bt884* transcripts was measured both by scintillation counting and densitometric scanning of autoradiographs with different exposure times. The relative transcriptional activity was determined by correcting the values for the UTP contents of the hybridizing RNA.

The cry*-coding region carries cryptic RNA processing sites*

In a next set of experiments we looked for unexpected processing events of the *cry* precursor mRNA which could interfere with export to the cytoplasm. To this end, the molecular structure of polyadenylated *cry* transcripts present in mesophyll protoplasts of N28-220 was studied. This mRNA pool should include the cytoplasmic *cry* mRNAs as well as the nuclear *cry* transcripts that have just undergone 3′ end formation. Total RNA of N28-220 protoplasts was extracted and cDNA was made using an oligo-d(T) primer. This cDNA was used in PCR reactions to amplify overlapping regions of the *cry* transcript to verify their integrity (Fig. 3a). Primers amplifying the region −8 to 672, and 539 to 1302 relative to the translation start site, yielded PCR products of the expected size only. Instead, amplification of the region 1264 to 1909 yielded in addition to the expected 645 bp product, smaller sized products of 437 and 298 bp, respectively (Fig. 3A). Sequence analysis showed that the 437 and 298 bp amplification products resulted from deletions of 208 nt, and 139 plus 208 nt, respectively, in the *bt884* transcript. The 5′ and 3′ internal and external borders of the missing regions are in both cases in agreement with RNA consensus splice sites (Fig. 3B). This finding indicates that a fraction of the *cry* precursor mRNA pool is spliced in a cryptic fashion. The 139 and 208 nt introns were named intron 1 and 2, respectively. In a next step, the *cry* transcript was screened for the occurrence of additional introns by amplifying it with oligos that prime at nt 107 and 1978. To increase the sensitivity, the PCR reaction mixture was blotted after gel separation and hybridised with a radioactively labelled *cry* probe. This revealed the amplification of a 0.6 and 0.7 kb product when separated on agarose gel (Fig. 3C). After purification of the individual bands followed by nested PCR, both bands gave rise to a single 0.7 kb product. The 0.6 kb product probably resulted from an imperfect duplex formation (see Materials and methods) [12, 46]. Sequence analysis revealed that the 0.7 kb PCR product had a size of 735 bp and contained a deletion of 929 and of 208 nt relative to the *bt884* sequence. The latter deletion corresponded with intron 2, whereas the 929 nt deletion uncovered the presence of a third cryptic intron. Interestingly, this intron uses the same 3′ splice site as intron 1. Taken together, these data show that the *cry*-coding region carries at least three cryptic introns (Fig. 3D).

Modification of the region 700–1000 and splice sites enhances cry *IA(b) expression*

A 5′ and 3′ deletion analysis of the *bt860* gene [42] using transiently transformed tobacco leaf protoplasts, revealed that the sequence between nucleotides 785 to 1285 interferes with *cry* mRNA accumulation (data not shown). To investigate whether this region is important for the low

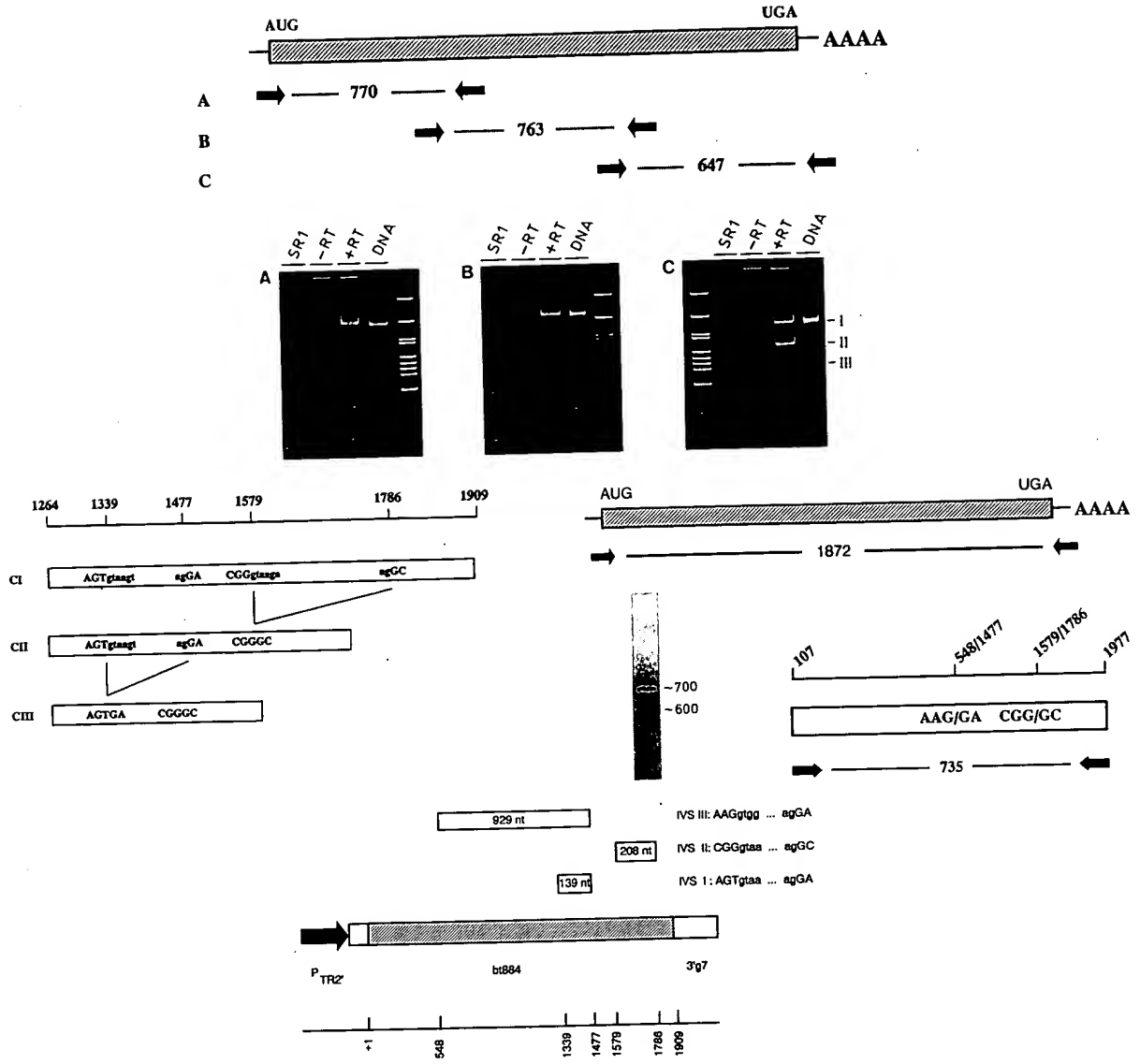

Fig. 3. Identification of cryptic RNA processing sites in the *bt884*-coding region. A (top). Search for splicing events in the 5′, central and 3′ region of the *bt884* transcript using RT-PCR. The primer sets used to amplify the regions A, B and C are indicated. Samples were separated on 8% polyacrylamide gel. SR1, RNA of untransformed SR1; −/+ RT, without or with reverse transcriptase; DNA, plasmid DNA positive PCR control. Size marker: pSTNV *Hinf* I. B (middle left). Sequence of the PCR products CI, CII and CIII. Sequences absent in the CII and CIII PCR products are written in lower-case letters. The nucleotide numbering is relative to the translation start site. C (middle right). RT-PCR of the entire *bt884*-coding region. The location of the primers is indicated. After separation on agarose gel the PCR products were visualized by Southern blotting using *bt884* DNA as a probe. Nucleotide numbering is relative to the AUG. D (bottom). Schematic overview of the three cryptic introns in the *bt884*-coding region. The 5′ and 3′ splice sites of the three introns are shown. IVS, intervening sequence.

nucleo-cytoplasmic *cry* mRNA flow, the expression of a *bt884*-derived gene modified at this position was studied. The region was altered by 63 translationally neutral substitutions between nucleotides 680 and 991 (Fig. 4A). The modified region was introduced into the *bt884* gene, yielding IA(b)-5. In addition it was introduced into the *bt884* derivative IA(b)-4, that carries a consensus

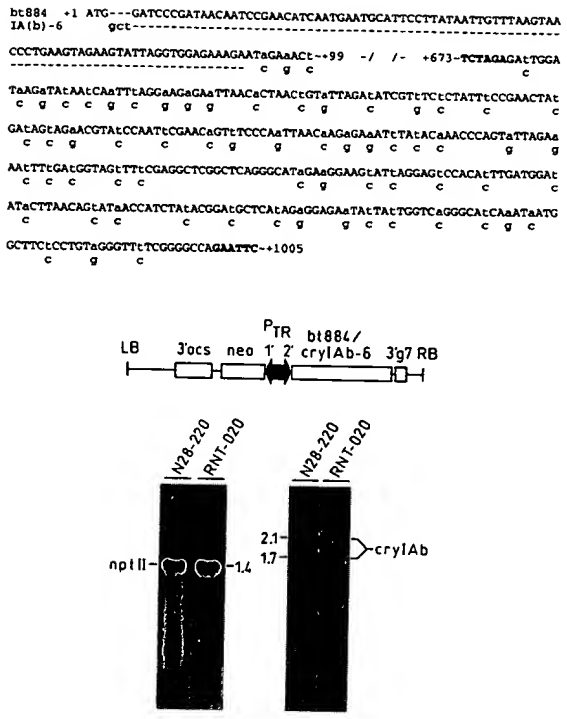

Fig. 4. Modification of the region 700–1000 results in improved *cry* expression. A (top). The configuration of the T-DNAs used to compare IA(b)-6 and *bt884* expression are schematically presented. Nucleotides differences between the *bt884* and IA(b)-6 coding regions are shown. Dotted line refers to nucleotides absent in either gene. Nucleotides written in bold refer to the restrictions sites used to introduce the modified pause site region. B (bottom). Comparison of the transcript levels directed by IA(b)-6 and *bt884* genes in mesophyll protoplasts prepared from stably transformed SR1 tobacco. The IA(b)-6 gene in RNT-020 directed ca. 5 pg/μg total RNA whereas the *bt884* gene in N28-220 directed less than 0.1 pg/μg total RNA. Abundances were determined using a dilution series of cold IA(b) and *neo* transcript. In RNT-020 a hybridizing RNA species of 1700 nucleotides accumulated in addition to the expected 2100 nt transcript. The 1700 nt RNA does not result from cryptic splicing (manuscript in preparation).

translation start context [23] and lacks the N-terminal 28 codons which are not necessary for insecticidal activity [20], thus yielding the IA(b)-6 gene. The resulting IA(b)-4, -5 and -6 genes were stably introduced into tobacco by T-DNA transfer and the IA(b) transcript levels were compared to that in N28-220. Interestingly, the IA(b)-4 and -5 genes typically directed a 10- to 20-fold higher IA(b) mRNA level than the *bt884* gene in N28-220 (data not shown). Figure 4B shows the mRNA level encoded by the IA(b)-6 gene in RNT-020 which is over 50-fold higher than that encoded by the *bt884* gene in N28-220. This suggests that the combination of the modifications as present in IA(b)-6 negatively cooperate with each other [18]. Transcript stability assays showed that the cytoplasmic stability of the IA(b)-6 transcript is similar to that of *bt884* (data not shown) which implies that the modifications in IA(b)-6 increase expression by enhancing the nucleo-cytoplasmic mRNA flow.

To determine whether cryptic splicing contributed to the low nucleo-cytoplasmic flow of full-size *bt884* mRNA the expression of IA(b)-6 genes with modified 5′ splice sites was studied. The 5′ splice sites of the introns 1 and 2 were modified by translationally neutral nucleotide changes and introduced into the IA(b)-6 gene. The expression of the modified genes was tested in transiently transformed tobacco leaf protoplasts and revealed that genes carrying a modification of the 5′ splice site of intron 2 either alone or in combination with the modified intron 1 5′ splice site, had a significant effect on transcript accumulation (Fig. 5A). The relative increase in expression was several-fold and depending on the experiment varied between 4- and 20-fold. An RT-PCR analysis of the transcripts revealed that the splice site modifications indeed abolished splicing (Fig. 5B). Taken together, this implies that cryptic splicing of the *cry* transcript has a significant negative effect on expression of the *bt884* gene.

Discussion

The cytoplasmic transcript level specified by chimaeric *cry* plant genes in a number of different plant species ranges without exception between undetectable and a few transcripts per cell. As the transgenes are flanked by regulatory signals which in combination with other coding sequences usually support efficient mRNA accumulation, it is believed that the *cry*-coding region somehow, in a cryptic fashion, negatively controls its expression in plants.

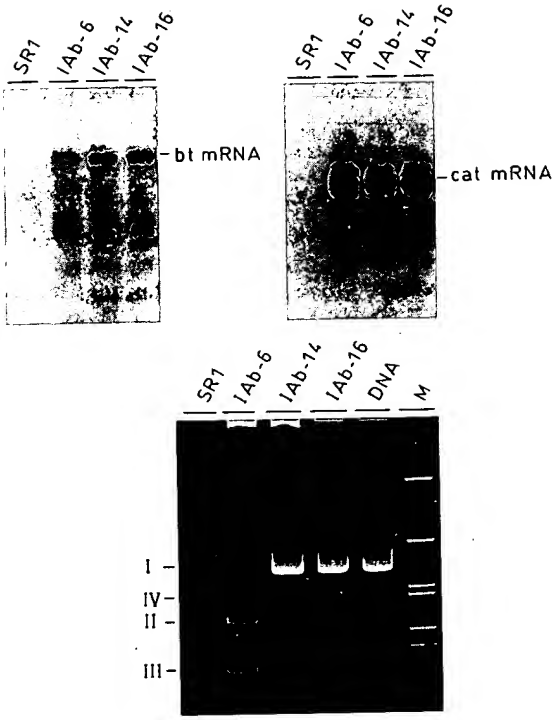

Fig. 5. Modification of the 5′ splice site of intron 2 improves *cry* expression. The 5′ splice sites of intron 1 and 2 were changed by translationally neutral substitutions (see Materials and methods). The modified 5′ splice sites were introduced into IA(b)-6 yielding IA(b)-14 that contains the modified 5′ splice site of intron 2, and IA(b)-16 that contains the modified 5′ splice sites of intron 1 and 2. A (top). RNA levels encoded by chimaeric IA(b)-6, -14 and -16 genes were compared in transiently transformed tobacco SR1 leaf protoplasts four hours after DNA delivery. The IA(b) genes were codelivered with a chimaeric *cat* reporter gene present on the same plasmid. Northerns contained a dilution series of *in vitro* transcribed *cry* IAb and *cat* RNA to quantify mRNA levels (not shown). B (bottom). RT-PCR of the RNA encoded by IA(b)-6, -14 and -16 genes shown in A. PCR products were separated on a 8% polyacrylamide gel. Product I is obtained with all genes and corresponds to full-size mRNA. Products II and III are obtained only with IA(b)-6 mRNA and correspond to removal of intron 2 and intron 1 and 2 respectively (see also Figs. 3A and 3B. Product IV is obtained exclusively with RNA encoded by IA(b)-14 and represents RNA from which intron 1 has been removed. SR1, RNA of untransformed SR1; DNA, plasmid DNA positive PCR control; M, pSTNV *Hinf* I size marker.

We studied several aspects of the *cry* IA(b) expression pathway to unmask the regulatory features of the coding region. In a first instance, we wished to determine in which cellular compartment expression is impaired. To this end, we determined the transcript level of the *bt884* gene in transgenic tobacco and its turnover in transiently and stably transformed tobacco. Integration of the data revealed that the *bt884* transcript is relatively stable and that the extremely low cytoplasmic *bt884* transcript level is primarily due to a restricted transcript flow to the cytoplasm. A similar observation was made for the *cry* IB, *cry* IC and *cry* IE genes (data not shown). This implies that the *cry* I-coding regions interfere with their own expression somewhere between the process of transcription initiation and nuclear export of the mature transcript. However, as the different *cry*-coding regions share only a moderate homology at the nucleotide level, the actual interfering mechanism may differ for each of them.

The relatively high stability of the *cry* I transcripts was somehow unexpected since the *cry* I coding regions display several features which may cause cytoplasmic mRNA instability. First, the AU/GC ratio deviates significantly from values found for plant coding regions and well expressed reporter genes as *nptII*, *bar*, *gus* and *cat*. A plant coding region typically has an AU content of about 40 to 50%, whereas the *cry* I-coding regions have an AU content of 60 to 64%, exceeding in some regions 70%. Regions carrying AU stretches have been implicated in the rapid turnover of some mRNA species in mammals [39, 44]. Furthermore, AU motifs which have been associated with specific mRNA decay pathways in mammals [1, 25, 29], appear at several positions in the coding sequence. Secondly, the codon usage of the *cry*-coding sequence is very unlike the preferred plant codon usage [31]. At several sites series of unfavourable codons are clustered, which might interfere with efficient translation elongation [45] and thus render the mRNA unstable because of an increased exposure of internal regions of the transcript to cytoplasmic RNase activities [17, 19, 33, 35, 38]. In conclusion, the stability of the *cry* transcripts exemplifies that AU-rich mRNA turnover signals are recognized differently within the cell [37] and that codon usage may have differential effects on mRNA stability.

Replacement of aberrant *cry* codons by preferred codons and an increase in the GC contents of the *cry* IA (b) coding sequence resulted in a five fold increase in the transcript level and a 10-fold increase of protein synthesis per mRNA [35]. These data show that codon usage and/or GC content indeed affect *CRY* protein synthesis in plants. The increase in mRNA level is not necessarily linked to an improved translation efficiency and/or mRNA stability, but could be due to improved nuclear processing and/or transport [40]. Noteworthy is the observation [35] that a directed modification about 250 nucleotides downstream of the start of translation significantly improves expression. This in fact confirms that the inhibitory mechanism is more complex than just inefficient translation due to an overall unfavourable codon usage.

As our data clearly indicated that *cry* expression is negatively controlled at the nuclear level, we tested whether the promoter of the chimaeric *bt884* transgene in tobacco was still active. The nuclear run-on analysis strongly indicates that the *bt884*-coding region does not significantly interfere with the frequency of transcription initiation, indicating that other processes were involved.

Studying the nuclear processing of the *bt884* precursor mRNA, we identified three functional introns in the *bt884* transcript. This was surprising because spliced *bt884* mRNAs do not accumulate to high levels in the cytoplasm. The only plausible explanation for this observation is that the spliced transcripts are inefficiently made and/or cytoplasmically unstable. Chimaeric IA(b) genes in which the 5′ splice site of intron 2 was mutagenised, directed significantly higher levels of full-length mRNA in transiently transformed tobacco mesophyll protoplasts. This implies that a large fraction of the *bt884* precursor mRNA enters into the splicing pathway and that cryptic splicing of the *bt884* mRNA either directly or indirectly interferes with transcript accumulation. We are currently investigating whether cryptic splicing causes an increased nuclear RNA turnover and/or affects cytoplasmic mRNA stability [4, 16], for instance by causing translational frameshifts. Furthermore, we are looking at the expression pathway of other *cry* genes to understand whether entry in the splicing pathway is a common event that interferes with *cry* gene expression in plants.

Acknowledgements

We thank Coby Slagter for technical assistance, Xavier Danthinne for providing plasmid pXD324, and Karel Spruyt, Vera Vermaercke and Christiane Germonprez for the illustrations.

References

1. Bandyopadhyay R, Coutts M, Krowczynska A, Brawerman G: Nuclease activity associated with mammalian mRNA in its native state: possible basis for selectivity in mRNA decay. Mol Cell Biol 10: 2060–2069 (1990).
2. Bossé M, Masson L, Brousseau R: Nucleotide sequence of a novel crystal protein gene isolated from *Bacillus thuringiensis* subspecies *kenyae*. Nucl Acids Res 18: 7443 (1990).
3. Brizzard B, Whiteley H: Nucleotide sequence of an additional crystal protein gene cloned from *Bacillus thuringiensis* subsp. *Thuringiensis*. Nucl Acids Res 16: 2723 (1988).
4. Burgess SM, Guthrie C: A mechanism to enhance mRNA splicing fidelity: The RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell 73: 1377–1391 (1993).
5. Celano P, Berchtold C, Casero R: A simplification of the nuclear run-off transcription assay. BioTechniques 7: 942–943 (1989).
6. Chen Z, Harless M, Wright D, Kellems R: Identification and characterisation of transcriptional arrest sites in exon1 of the human adenosine deaminase gene. Mol Cell Biol 10: 4555–4564 (1990).
7. Cornelissen M: Nuclear and cytoplasmic sites for anti-sense control. Nucl Acids Res 17: 7203–7209 (1989).
8. Cornelissen M, Vandewiele M: Both RNA level and translation efficiency are reduced by anti-sense RNA in transgenic tobacco. Nucl Acids Res 17: 833–843 (1989).
9. Cox S, Goldberg R: Analysis of plant gene expression, In: Shaw C (ed) Plant Molecular Biology: A Practical Approach, pp. 1–35. IRL Press (1989).
10. Danthinne X, Seurinck J, Meulewaeter F, Van Montagu M, Cornelissen M: The 3′ untranslated region of satellite tobacco necrosis virus RNA stimulates translation *in vitro*. Mol Cell Biol 13: 3340–3349 (1993).
11. Denecke J, Gosselé V, Botterman J, Cornelissen M:

Quantitative analysis of transiently expressed genes in plant cells. Meth Mol Cell Biol 1: 19–26 (1989).
12. Foley PF, Leonard MW, Engel JD: Quantitation of RNA using the polymerase chain reaction. Trends Genet 9: 380–385 (1993).
13. Gallie D: The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Devel 5: 2108–2116 (1991).
14. Gallie D, Lucas W, Walbot V: Visualising mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. Plant Cell 1: 301–311 (1989).
15. Gallie D, Walbot V: RNA pseudoknot domain of tobacco mosaic virus can functionally substitute for a poly(A) tail in plant and animal cells. Genes Devel 4: 1149–1157 (1990).
16. Green PJ: Control of mRNA stability in higher plants. Plant Physiol 102: 1065–1070 (1993).
17. Herrick D, Parker R, Jacobson A: Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. Mol Cell Biol 10: 2269–2284 (1990).
18. Herschlag D, Johnson F: Synergism in transcriptional activation: a kinetic view. Genes Devel 7: 173–179 (1993).
19. Hoekema A, Kastelein R, Vasser M, Boer H: Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol Cell Biol 7: 2914–2924 (1987).
20. Höfte H, Whitely H: Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol Rev 53: 242–255 (1989).
21. Honée G, van der Salm T, Visser B: Nucleotide sequence of crystal protein gene isolated from *B. thuringiensis* subspecies *entomocidus* 60.5 coding for a toxin highly active against Spodoptera species. Nucl Acids Res 16: 6240 (1988).
22. Horton R, Hunt H, Ho S, Pullen J, Pease L: Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77: 61–68 (1989).
23. Joshi C: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl Acids Res 15: 6643–6653 (1987).
24. Kawasaki ES: Amplification of RNA, In: Innes MG, Gelfand D, Sninski J, White T, (eds) PCR protocols: A guide to methods and applications, pp. 21–27. Academic Press (1990).
25. Kruys V, Marnix O, Shaw G, Deschamps J, Hue G: Translational blockade imposed by cytokine-derived UA-rich sequences. Science 245: 852–855 (1989).
26. Laird P, Zomerdijk J, de Korte D, Borst P: *In vivo* labelling of intermediates in the discontinuous synthesis of mRNAs in *Trypanosoma brucei*. EMBO J 6: 1055–1062 (1987).
27. Maas C, Laufs J, Grant S, Korfhage C, Werr W: The combination of a novel stimulatory element in the first exon of the maize *Shrunken-1* gene with the following intron 1 enhances reporter gene expression up to 1000-fold. Plant Mol Biol 16: 199–207 (1991).
28. Maliga P, Breznovitz A, Marton L: Streptomycin resistant plants from callus culture of haploid tobacco. Nature 244: 29–30 (1973).
29. Malter J: Identification of an AUUUA-specific messenger RNA binding protein. Science 246: 664–666 (1989).
30. Maxam A, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. Meth Enzymol 65: 499–560 (1980).
31. Murray E, Lotzer J, Eberle M: Codon usage in plant genes. Nucl Acids Res 17: 477–498 (1989).
32. Murray E, Rocheleau T, Eberle M, Stock C, Sekar V, Adang M: Analysis of unstable RNA transcripts of insecticidal crystal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. Plant Mol Biol 16: 1035–1050 (1991).
33. Neill J, Litts J, Anderson O, Greene F, Stiles J: Expression of a wheat gliadin gene in *Saccharomyces cerevisiae*. Gene 55: 303–317 (1987).
34. Perlak F, Deaton R, Armstrong T, Fuchs R, Sims R, Greenplate J, Fischhoff D: Insect resistant cotton. Bio/technology 8: 939–943 (1990).
35. Perlak F, Fuchs R, Dean D, McPherson S, Fischhoff D: Modification of the coding sequence enhances plant expression of insect control protein genes. Proc Natl Acad Sci USA 88: 3324–3328 (1991).
36. Sanger F, Nicklen S, Coulson A: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
37. Schuler G, Cole M: GM-CSF and oncogene stabilities are independently regulated in *trans* in a mouse monocytic tumor. Cell 55: 1115–1122 (1988).
38. Sharp P, Devine K: Codon usage and gene expression in *Dictyostelium discodeum*: highly expressed genes do 'prefer' optimal codons. Nucl Acids Res 17: 5029–5039 (1989).
39. Shaw G, Kamen R: A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46: 659–667 (1986).
40. Urlaub G, Mitchell P, Ciudad C, Chasin L: Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. Mol Cell Biol 9: 2868–2880 (1989).
41. Vaeck M, Reynaerts A, Hofte H: Protein engineering in plants: expression of Bacillus thuringiensis insecticidal protein genes. Cell Cult Somatic Genet Plants 6: 425–439 (1989).
42. Vaeck M, Reynaerts A, Hofte H, Jansens S, De Beuckeleer M, Dean C, Zabeau M, Van Montagu M, Leemans L: Transgenic plants protected from insect attack. Nature 327: 33–37 (1987).
43. Yanisch-Perron C, Vieira J, Messing J: Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119 (1985).
44. Wilson T, Treisman R: Removal of poly(A) and conse-

quent degradation of c-*fos* mRNA facilitated by 3′ AU-rich sequences. Nature 336: 396–399 (1988).
45. Wolin S, Walter P: Ribosome pausing and stacking during translation of a eucaryotic mRNA. EMBO J 7: 3559–3569 (1988).
46. Zacharias DA, Garamszegi N, Strehler EE: Characterization of persistent artifacts resulting from RT-PCR of alternative spliced mRNAs. BioTechniques 17: 652–655 (1994).